JOURNAL OF BACTERIOLOGY, May 2004, p. 3274–3277 Vol. 186, No. 10
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.10.3274–3277.2004

−1 Frameshifting at a CGA AAG Hexanucleotide Site Is Required for Transposition of Insertion Sequence IS*1222*

Nina Mejlhede,[1,2]† Patricia Licznar,[1] Marie-Françoise Prère,[1] Norma M. Wills,[2] Raymond F. Gesteland,[2] John F. Atkins,[2] and Olivier Fayet[1]*

Laboratoire de Microbiologie et Génétique Moléculaire, UMR5100 Centre National de la Recherche Scientifique et Université Paul Sabatier, Toulouse 31062-cedex, France,[1] and Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112-5330[2]

Received 29 October 2003/Accepted 2 February 2004

The discovery of programmed −1 frameshifting at the hexanucleotide shift site CGA_AAG, in addition to the classical X_XXY_YYZ heptanucleotide shift sequences, prompted a search for instances among eubacterial insertion sequence elements. IS*1222* has a CGA_AAG shift site. A genetic analysis revealed that frameshifting at this site is required for transposition.

Signals in some mRNAs direct a proportion of translating ribosomes to shift the reading frame, thereby providing regulatory or novel product possibilities not achievable by standard decoding (1, 3). Nearly all cases of −1 frameshifting occur by tandem slippage of tRNA anticodons on heptanucleotide shift sites (7). One of the few exceptions with re-pairing of only a single tRNA is the −1 frameshifting in decoding the *Bacillus subtilis* cytidine deaminase gene (*cdd*). The realignment is from AAG to AAA in the sequence CGA_AAG, with importance for the codon 5′ of AAG being CGA (11). The efficiency of *cdd* frameshifting is 15%, and as with the −1 frameshifting utilized in decoding *Escherichia coli dnaX* (7), *cdd* frameshifting is stimulated by an internal Shine-Dalgarno (SD) sequence, GGAGG, 10 bases 5′ of the shift site (11). A role for *cdd* frameshifting in influencing expression of an overlapping downstream gene has been considered (11) but is not supported by comparative sequence analysis (5). To determine if CGA_AAG-based frameshifting is required for gene expression, we searched for other potential −1 frameshift signals which use this motif.

Bacterial insertion sequences (IS) constitute the richest source of sequences where −1 frameshifting is either known or suspected to be utilized for gene expression (4). One of them, IS*1222* from *Rahnella aquatilis* (16), has CGA_AAG as its suspected frameshift site in the 34-nucleotide overlap between two consecutive genes, *orfA* and *orfB* (Fig. 1). Eight nucleotides 5′ of it there is a potential internal SD sequence, AG GUGG. For some related IS elements, a frameshift-generated fusion protein, OrfAB, is essential for transposition (IS*911* [12], IS*3* [15], and IS*150* [17]). A −1 frameshift event at the CGA_AAG motif of IS*1222* would lead to synthesis of a very similar protein.

All tandem slippage frameshift sites known to be utilized for gene expression have an RNA structural element 3′ of the shift site that increases the proportion of ribosomes that change frame. Six nucleotides 3′ of the IS*1222* CGA_AAG motif there is a potential 9-bp stem topped with a structured region (Fig. 1A). It is reminiscent of the only branched stem-loop structure known to be involved in stimulation of frameshifting, which is involved in decoding IS*911* (14). In addition, the more distal UCACA sequence could pair with UGUGA as indicated in Fig. 1A (16), leading to the formation of a pseudoknot resembling the only bacterial frameshift stimulatory pseudoknot known, that occurring downstream of the A_AAG shift site of IS*3* (15). In the stem-loop structure, there are imperfect SD-like sequences followed by three consecutive codons, GUG_AUG_UUG, which could be potential initiators for OrfB synthesis.

Identification of the shift site and frameshifting stimulators. The IS*1222* putative shift region was cloned between phage T7 gene *10* and *E. coli lacZ* in the vector pOFX302 (Fig. 1B) (14); to facilitate protein purification, the same region was also cloned into vector pGMH (5). Gene *10*, or *gst*, and the distal part of *orfA* are in the same 0 phase, and the proximal part of *orfB* is in frame with *lacZ*, or *malE* (−1 phase).

Pulse-labeling experiments show that decoding the IS*1222* cassette results in 4.4% frameshifting as well as initiation of OrfB synthesis (6%) (Fig. 1C). When the CGA_AAG hexamer was changed to CG**C**_AAG, CGA_AAA, CGA_AA**C**, or CG**C**_AAA, frameshifting was not detected, suggesting that CGA_AAG is the IS*1222* shift site. Direct evidence was provided by mass spectrometric analysis of the protease-cleaved Gst-MalE frameshift product: the mass of the observed product corresponds exactly to the expected 46,658-Da protein (Fig. 1D). Next, the 5′ internal SD sequence was weakened (to UCCGUCG) with a resultant twofold decrease in frameshifting. When it was strengthened (from UCAGGUGG to UAAG GAGG), there was a 1.3-fold increase in frameshifting. However, previous results have shown that eight nucleotides is a suboptimal spacing for −1 frameshifting stimulatory SD sequences (8). This is perhaps why the IS*1222* SD frameshifting stimulator is fivefold less efficient than its *cdd* counterpart (11).

The putative 3′ structure was also investigated by mutagen-

* Corresponding author. Mailing address: Laboratoire de Microbiologie et Génétique Moléculaire, UMR5100 Centre National de la Recherche Scientifique et Université Paul Sabatier, 118 route de Narbonne, Toulouse 31062-cedex, France. Phone: 33 5 61 33 58 75. Fax: 33 5 61 33 58 86. E-mail: Olivier.Fayet@ibcg.biotoul.fr.

† Present address: Riso National Laboratory, Plant Research Department, DK-4000 Roskilde, Denmark.

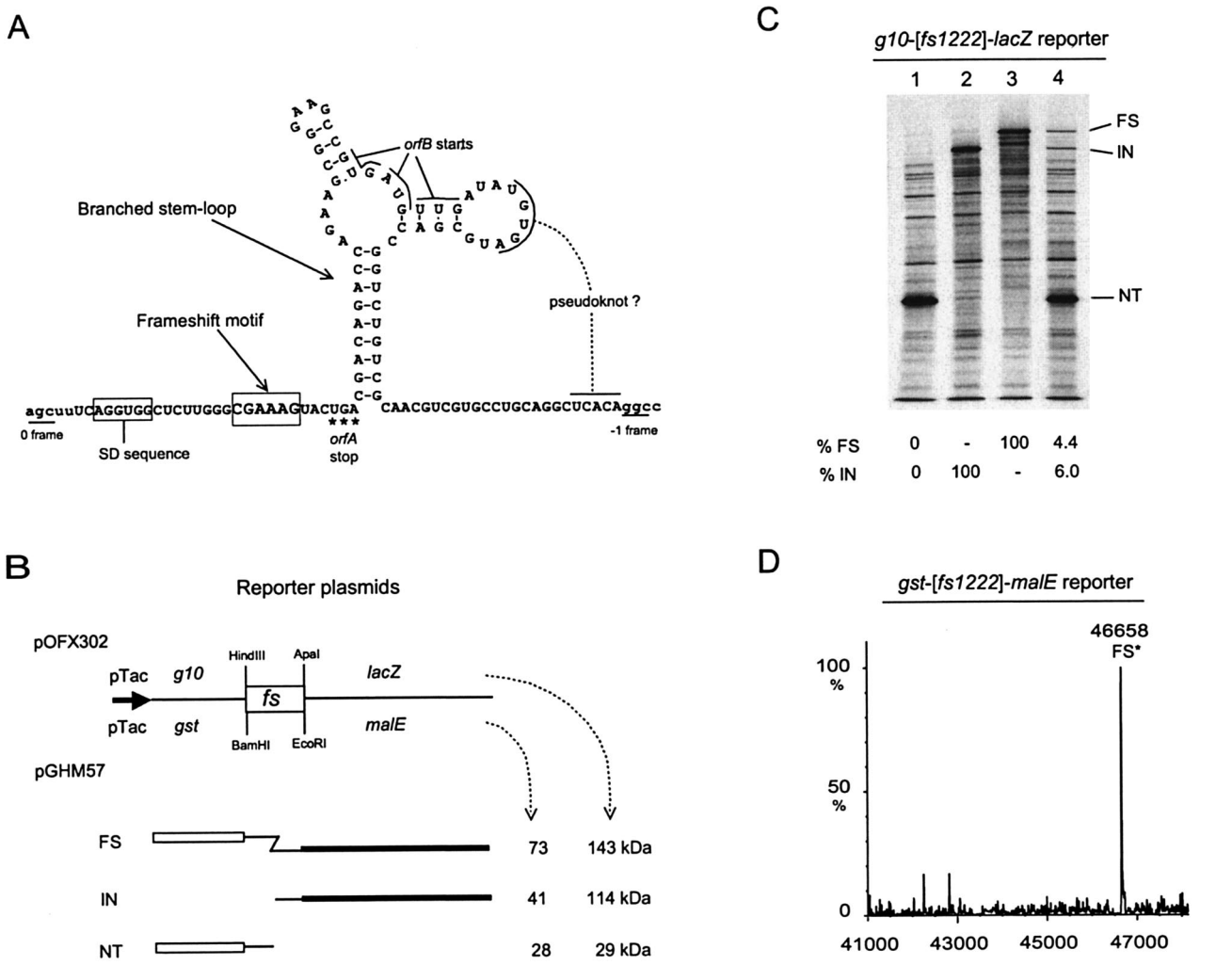

FIG. 1. Functional analysis of the IS*1222* frameshift region. (A) Signals and predicted branched stem-loop structure in the IS*1222* frameshift region. (B) The IS*1222* frameshift region shown in panel A was cloned into plasmids pOFX302 (14) and pGHM (i.e., the GM1 derivative with a C-terminal His tag and a PreScission protease site described previously [6]), by using the indicated restriction sites. In the pOFX302 construct, normal translation of gene *10* should end at the indicated UGA stop codon in the cloned IS*1222* region to give a 274-amino-acid product (NT). Frameshifting on CGA_AAG should lead to synthesis of a 1,325-amino-acid protein (FS). Independent initiation at the GUG codon in the IS*1222* *orfB* frame should give a 1,041-amino-acid product (IN). (C) Translation products from the plasmid-borne gene *10-lacZ* region were detected by in vivo [^{35}S]methionine pulse-labeling polyacryamide gel electrophoresis and quantitated with a Fuji PhosphoImager (2, 14). The frameshifting and initiation capacities of the IS*1222* recoding region (lane 4) were estimated by comparison with reference strains (lanes 1 to 3). Lane 1 corresponds to the labeling of the vector-containing strain used for background correction. Lanes 2 and 3 respectively correspond to constructs expressing at the maximum level the LacZ (100% IN) and G10-LacZ (100% FS) products (plasmids pOFX302-0 and pOFX302-4, described in reference 14). (D) Mass spectrum of a proteolytic fragment (FS*) from the frameshift product purified from a strain containing the IS*1222* region cloned into plasmid pGMH. The 73-kDa frameshift product was purified, digested with PreScission protease (Amersham), and analyzed as previously described (6). The observed mass of the FS* product (46,658 Da) is exactly that expected for the proteolytic fragment derived from a Gst-MalE fusion protein generated through −1 frameshifting, from AAG to the overlapping AAA codon, on the CGA_AAG motif.

esis (Fig. 2). Deletion of its 3′ half abolishes frameshifting (Fig. 2A). A deletion from the 3′ side, which leaves intact the branched stem-loop, is without notable effect, thus showing that the potential pseudoknot does not contribute significantly to frameshifting stimulation (Fig. 2B). Mutations that prevent formation of a full-length 9-bp stem give a threefold reduction in frameshifting, and compensatory changes restore wild-type levels (Fig. 2C).

Frameshifting is required for transposition. To study the role in transposition of the frameshift-generated OrfA-OrfB transframe product (OrfAB), a functional copy of IS*1222* was obtained from a clinical isolate of *R. aquatilis* by PCR amplification and cloning into *E. coli*, by using plasmid pAT153 as the vector. This IS*1222* variant displays 147 nucleotide changes (GenBank accession number AY528232) when compared to the original isolate (16); none of them affects the shift site or the two associated stimulators. It was found to have uninterrupted *orfA* and *orfB* genes: in terms of amino acids, 104 changes are neutral, 26 are conservative, and 6 are nonconservative. More importantly, this cloned variant proved to be

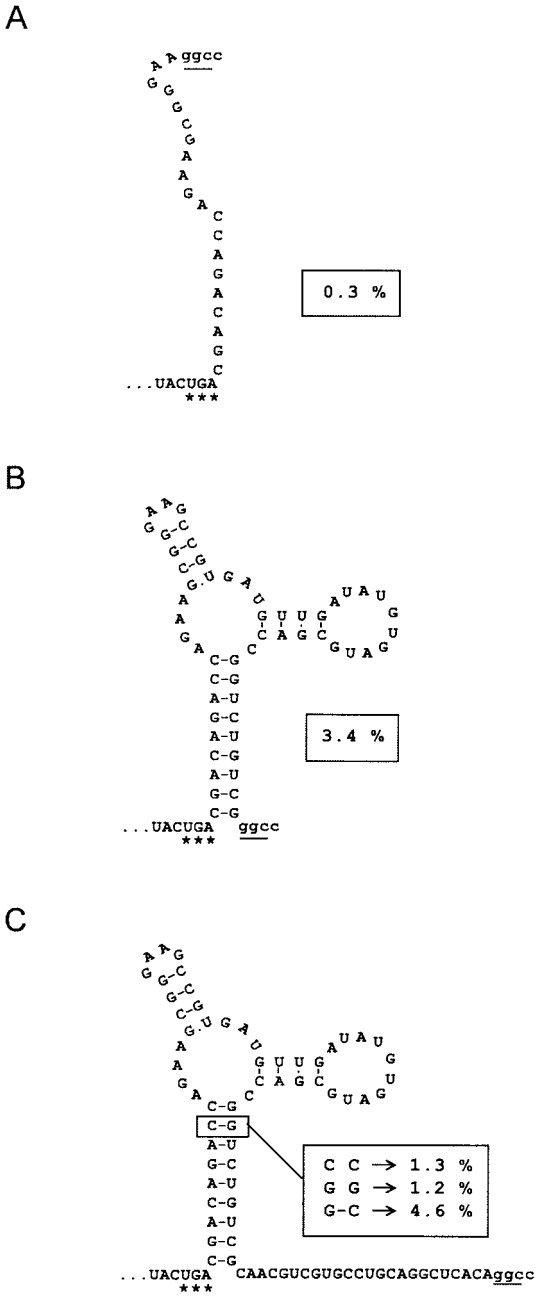

FIG. 2. Genetic analysis of the IS*1222* stimulatory stem-loop. (A) Deletion of the 3′ half of the structure. (B) Deletion of the region 3′ of the stem-loop removing the putative pseudoknot. (C) Set of three mutants in which stability of the bottom stem is either altered or restored. All mutants were generated by cloning oligonucleotides into plasmid pOFX302 and subsequent pulse-labeling analysis (2, 14).

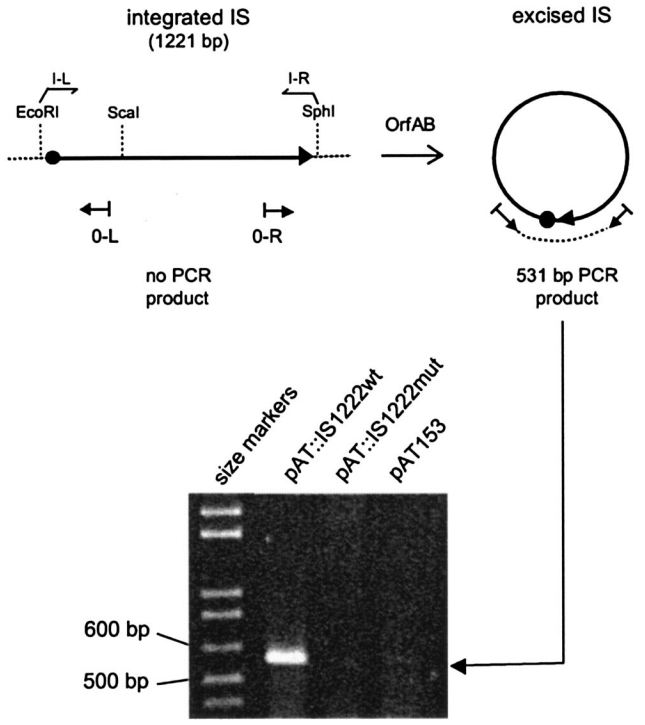

FIG. 3. Transposition activity in *E. coli* of a frameshift-proficient and a frameshift-deficient IS*1222* variant. An active IS*1222* variant was amplified from the genome of a clinical isolate of *R. aquatilis* and cloned into the EcoRI and SphI sites of plasmid pAT153, to give pAT::IS*1222*wt. Two oligonucleotides (I-L and I-R), corresponding to the left and right ends of the IS as previously published (accession number X78052) (16), were used for amplification. The frameshift motif was then changed to CG**C**_AA**A**, to prevent frameshifting, by using PCR mutagenesis and the plasmid EcoRI site and the IS ScaI site overlapping the CGA_AAG motif, thus giving plasmid pAT::IS*1222*mut (experimental details available upon request). Plasmid DNA was prepared from strains containing either pAT153 or its two IS-containing derivatives, linearized with PstI, and tested in PCRs containing two divergent IS oligonucleotides (O-L, ACCTCCATACC GCCATACTTCTTACGCCA, and O-R, TATCCGGCGACAATAA GAACCGATCAG). Authenticity of the 531-bp fragment obtained with pAT::IS*1222*wt was verified by sequencing.

active in transposition since it can promote its self-excision from the pAT153 plasmid to give an IS circle, a known intermediate in transposition (9, 13). This was shown by carrying out a PCR assay with two outward-oriented IS-specific oligonucleotides (Fig. 3). The presence of a 531-bp fragment indicates IS circle formation, hence transposition capacity (Fig. 3, lane pAT::IS*1222*wt). The CGA_AAG motif was changed to CG**C**_AA**A**, to abolish frameshifting without changing the amino acid sequence of the OrfA protein. This, in turn, eliminates IS circle formation (Fig. 3, lane pAT::IS*1222*mut). Thus, frameshifting on the CGA_AAG motif is essential for IS*1222* transposition.

Originality of the IS*1222* frameshift signal. The IS*1222* recoding cassette defines a new type of composite frameshift signal. It has the second known example of a CGA_AAG hexameric shift site at which only a single tRNA re-pairs; a recent study provides clues for its shiftiness and for that of other related motifs (10). Frameshifting is stimulated by an upstream SD sequence, as in *dnaX*, and more strongly by a downstream branched stem-loop previously known only for IS*911* frameshifting. The combination gives a moderate frameshifting efficiency, 4.4%, which is nevertheless enough to provide the amount of transframe protein required for transposition of IS*1222*. Analysis of this region emphasizes the modular nature of prokaryotic frameshift sites as well as the flexibility of their evolution. These features will have to be taken into ac-

count in searches to determine how frequently frameshifting is used in gene expression.

Nucleotide sequence accession number. The sequence of the entire IS*1222* variant was verified and deposited in GenBank under accession number AY528232.

We are very grateful to Isabelle Canal for expert technical help.

This work was supported by EMBO (short-term fellowship to N.M. for her work in Toulouse), by the Centre National de la Recherche Scientifique, by the Université Paul Sabatier of Toulouse, by a grant to O.F. (programme de recherche fondamentale en microbiologie et maladies infectieuses et parasitaires), by a Department of Energy grant to R.F.G. (DEFG03-99ER62732), and by an NIH grant to J.F.A. (R01-GM48152).

REFERENCES

1. **Baranov, P. V., R. F. Gesteland, and J. F. Atkins.** 2002. Recoding: translational bifurcations in gene expression. Gene **286:**187–201.
2. **Bertrand, C., M. F. Prère, R. F. Gesteland, J. F. Atkins, and O. Fayet.** 2002. Influence of the stacking potential of the base 3′ of tandem shift codons on −1 ribosomal frameshifting used for gene expression. RNA **8:**16–28.
3. **Brierley, I., and S. Pennell.** 2001. Structure and function of the stimulatory RNAs involved in programmed eukaryotic −1 frameshifting. Cold Spring Harbor Symp. Quant. Biol. **66:**233–248.
4. **Chandler, M., and O. Fayet.** 1993. Translational frameshifting in the control of transposition in bacteria. Mol. Microbiol. **7:**497–503.
5. **Gurvich, O. L., P. V. Baranov, J. Zhou, A. W. Hammer, R. F. Gesteland, and J. F. Atkins.** 2003. Sequences that direct significant levels of frameshifting are frequent in coding regions of *Escherichia coli.* EMBO J. **22:**5941–5950.
6. **Herr, A. J., C. C. Nelson, N. M. Wills, R. F. Gesteland, and J. F. Atkins.** 2001. Analysis of the roles of tRNA structure, ribosomal protein L9, and the bacteriophage T4 gene 60 bypassing signals during ribosome slippage on mRNA. J. Mol. Biol. **309:**1029–1048.
7. **Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus.** 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. Cell **55:**447–458.
8. **Larsen, B., N. M. Wills, R. F. Gesteland, and J. F. Atkins.** 1994. rRNA-mRNA base pairing stimulates a programmed −1 ribosomal frameshift. J. Bacteriol. **176:**6842–6851.
9. **Licznar, P., C. Bertrand, I. Canal, M. F. Prère, and O. Fayet.** 2003. Genetic variability of the frameshift region in IS*911* transposable elements from *Escherichia coli* clinical isolates. FEMS Microbiol. Lett. **218:**231–237.
10. **Licznar, P., N. Mejlhede, M. F. Prère, N. Wills, R. F. Gesteland, J. F. Atkins, and O. Fayet.** 2003. Programmed translational −1 frameshifting on hexanucleotide motifs and the wobble properties of tRNAs. EMBO J. **22:**4770–4778.
11. **Mejlhede, N., J. F. Atkins, and J. Neuhard.** 1999. Ribosomal frameshifting during decoding of *Bacillus subtilis cdd* occurs at the sequence CGA AAG. J. Bacteriol. **181:**2930–2937.
12. **Polard, P., M. F. Prère, M. Chandler, and O. Fayet.** 1991. Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS*911.* J. Mol. Biol. **222:**465–477.
13. **Polard, P., M. F. Prère, O. Fayet, and M. Chandler.** 1992. Transposase-induced excision and circularization of the bacterial insertion sequence IS*911.* EMBO J. **11:**5079–5090.
14. **Rettberg, C. C., M. F. Prère, R. F. Gesteland, J. F. Atkins, and O. Fayet.** 1999. A three-way junction and constituent stem-loops as the stimulator for programmed −1 frameshifting in bacterial insertion sequence IS*911.* J. Mol. Biol. **286:**1365–1378.
15. **Sekine, Y., N. Eisaki, and E. Ohtsubo.** 1994. Translational control in production of transposase and in transposition of insertion sequence IS*3.* J. Mol. Biol. **235:**1406–1420.
16. **Steibl, H., and F. Lewecke.** 1995. IS*1222*: analysis and distribution of a new insertion sequence in *Enterobacter agglomerans* 339. Gene **156:**37–42.
17. **Vogele, K., E. Schwartz, C. Welz, E. Schiltz, and B. Rak.** 1991. High-level ribosomal frameshifting directs the synthesis of IS*150* gene products. Nucleic Acids Res. **19:**4377–4385.